Advanced Science and Technology Letters
Vol.52 (IUrC 2014), pp.131-141
http://dx.doi.org/10.14257/astl.2014.52.23

Effective Feature Selection Method Using Hog-Family Feature for Human Detection

Kitae Bae[1], Libor Mesicek[2], and Hoon Ko[3]

[1] Department of New Media, Korean German Institute of Technology
661, Deungchon-Dong, Gangseo-Gu, Seoul, 157-033, Republic of Korea
ktbae@kgit.ac.kr
[2, 3] Department of Informatics, J. E. Purkinje University, Faculty of Science,
Ceske mladeze 8, Usti nad Labem, Czech Republic, 400-96
[2]libor.mesicek@gmail.com, [3]hoon.ko@ujep.cz

Abstract. Support Vector Machine (SVM) is one of powerful learning machine and has been applied to varying task with generally acceptable performance. The success of SVM for classification tasks in one domain is affected by features which represent the instance of specific class. Given the representative and discriminative features, SVM learning will give good generalization and consequently we can obtain good classifier. In this paper, we will assess the problem of feature choices for human detection tasks and measure the performance of each feature. Here we will consider HOG-family feature. We proposed the multi-scale HOG as a NEW family member in this feature group. We also combine SVM with Principal Component Analysis (PCA) to reduce dimension of features and enhance the evaluation speed while retaining most of discriminative feature vectors.

1 Introduction

Detecting human has attracted a lot of research interests in recent years, due to the drive from many emerging applications, such as perceptual interfaces, ubiquitous computing, and smart video surveillance [1]-[3]. Different applications are concerned with different image resolutions of the subjects, thus requiring different techniques. For example, in perceptual interfaces, the motions of the human body parts need to be determined for action recognition; thus this application require fairly high resolution for analyzing the articulated motion of the body parts. In contrast, in many video surveillance applications, such as robot vision for inspection since the human typically is associated with small region and low resolutions, the human needs to be treated as a non-rigid entity for detection, while detailed motion of the body parts is no longer the major focus here. This paper addresses effective feature selection in the detection problem. In computer vision field, human detection is considered a difficult problem, due to the variation of human appearances in image. The complexity of the problem is added with the cluttered and dynamic background. In this paper, we consider human detection by placing this problem as a classification problem, solving with a discrimi-

ISSN: 2287-1233 ASTL

native approach through the supervised learning. Being in this manner, the human detection requires two components: a set of features and a discriminative learning method. It is important to have features that can robustly represent the appearance of object in interest, i.e. human, while optimally separates human from others, e.g. windows, chairs, etc. To classify these features, we also need a strong discriminative learning algorithm, which offers proper generalization and bring a desirable separation line between object and non-object. As shown in [4], Histogram of Oriented Gradient (HOG) has successfully trained the Support Vector Machines (SVM) to detect human in images with an acceptable performance. This success, in one aspect, might be considered the fruits of the success of the feature to model the observed object as an instance of a specific class. Given a discriminative set of features, SVM will give a good generalization and consequently provides a well classifier. However, finding a representative feature to cover every variation of pose and appearance of human in image is considered too difficult. It is also interesting to use redundant features for human detection problem, and employs AdaBoost algorithm to learn important features, while avoiding the rest to be considered in the model [5]. The HOG, as shown in [6], one of the recent robust features set for human detection problem (one can refer to Region Covariance [5] as an alternative) that encodes the object's shape by using gradient structure, and then captures the object spatial information by grid quantization and local normalization. This procedure lets HOG committed the illumination invariance as its property. Regarding the two main components of human detection in supervised learning approach, the remainders of this paper are mainly discussed two issues. The first one is the evaluation of several variants of Histogram of Oriented Gradient [4], so-called HOG-family, as a robust features set for discriminating human from other objects. This includes the non-overlapped, dense [4], spatial pyramidal [6] and our new proposed HOG feature: multi-scale image pyramid HOG. The systematic comparison within HOG-family along various combinations with SVM will provide experimental validation of the most discriminative feature within HOG-family. The second issue is the discussion about speeding-up the performance of SVM evaluation, trained by the HOG-family, by firstly reducing the features dimension through Principal Component Analysis (PCA). This reduction is expected to affect the speed of testing phase of SVM, while preserving the classification quality.

This paper is organized as follows. In Section 2, we describe the HOG-family as features, along with SVM as the classifier. The incorporation of PCA for feature dimension reduction with SVM is presented afterwards. The data sets and the performance measure are outlined in Section 3, with comparison and analysis as closing notes. Finally, conclusion and outline for future research direction is drawn in Section 4.

2 Features and the classifier

In this section we briefly introduce the HOG as our discriminative feature for dealing with the human detection problem. Afterwards, we discuss the basic concept of SVM and how PCA are incorporated into SVM for feature dimensional reductions. The

SVM under PCA transformation is proven to be invariant, there for it is safe to improve SVM evaluation along PCA.

2.1 HOG Family

The utilization of orientation histogram as shape encoding descriptors has been used in [8] for hand gesture recognition. Later it was developed into robust local feature descriptors known as SIFT [9]. In SIFT, the features are computed at a sparse set of scale-invariant key points, rotated to align their dominant orientations and used individually. The HOG as proposed by [4,10], are computed in dense grids at a single scale without dominant orientation alignment. The grid position of the block implicitly encodes spatial position relative to the detection window in the final feature vector. In addition to dense HOG [4], the experiment is also performed to evaluate non-overlapped HOG as the basic HOG variant,

The HOG-family features are shown in Fig. 1 and briefly described as follows:

(1) Non-overlapped. This is the simplest HOG representation. Histogram of each cell is normalized with the block norm.

(2) Overlapped (dense). This is similar to the implementation in [4], and it uses the overlapped blocks as the basic descriptor.

(3) Spatial pyramid. The pyramid of HOG is built by using different cells of each pyramid level and then the concatenated histogram is composed of feature data in each level.

(4) Multi-scale. The image pyramid levels are created from the image and the histogram is calculated at each level. By then, the feature vector is built by the concatenation of histogram from different levels. In this implementation, we combine two different level of image pyramid to build the HOG feature.

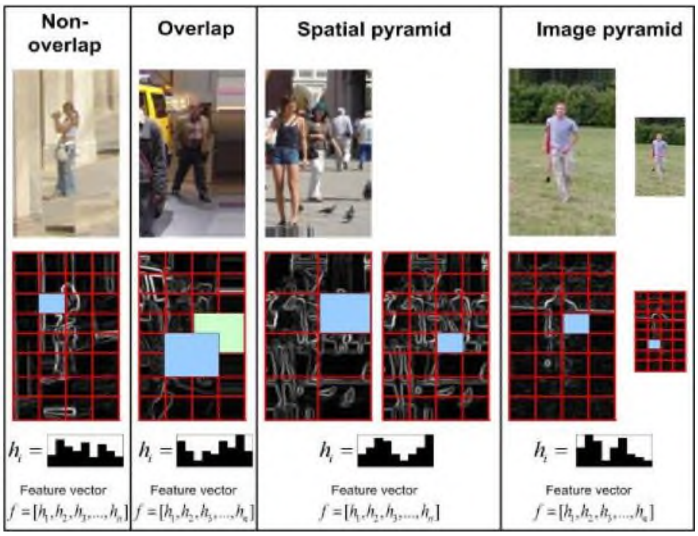

Non-overlap
Overlap
Spatial pyramid
Image pyramid
$h_i =$
Feature vector
$f = [h_1, h_2, h_3, \ldots, h_n]$
$h_i =$
Feature vector
$f = [h_1, h_2, h_3, \ldots, h_q]$
$h_i =$
Feature vector
$f = [h_1, h_2, h_3, \ldots, h_n]$
$h_i =$
Feature vector
$f = [h_1, h_2, h_3, \ldots, h_q]$

Fig. 1. HOG-Family features for human detection

2.2 Building HOG from image

Let I denote an image of width and represent the pixel intensity in corresponding position , then the HOG descriptor can be computed by the following steps:

(1) Perform image enhancement through gamma correction.
(2) From image , we compute image gradient in horizontal and vertical directions by one dimensional entered mask . By then, we have following formulations for image gradient in each point:

$$g_x(x,y) = I(x+1,y) - I(x-1,y)$$
$$g_y(x,y) = I(x+1,y) - I(x-1,y) \qquad (1)$$

where $g_x(x,y)$ and $g_y(x,y)$ denote the horizontal and vertical components of the image gradient respectively.
(3) The magnitude and the orientation of image data from its gradient can be calculated as follows:

$$m(x,y) = \sqrt{g_x(x,y)^2 + g_x(x,y)^2}$$
$$\theta(x,y) = \tan^{-1}(g_x(x,y)/g_x(x,y)) \qquad (2)$$

The orientation $\theta(x,y)$ has values in $[0,2\pi]$ range.
(4) We divide the image into $S_c X S_c$ non-overlapping cells. For each cells, we quantize the orientation $\theta(x,y)$ s_t for all pixels into orientation bins weighted by its magnitude .
(5) The feature is normalized by the sum of blocks. In all HOG variant, we use a magnitude of a block of cells to normalize each of cells. By blocks normalization S_c we S_c capture the information in the surroundings cells. We divide the image into -by- non-overlapping cells. For each cells, we quantize the orientation $\theta(x,y)$ for all pixels into orientation bins weighted by its magnitude .
(6) The histogram from each cell is formed into one feature vector which depends on the configuration required for each HOG variant as shown in Figure 1.

The output of this procedure per each image is 1-by-N feature contains normalized both image magnitude and orientation, which is considerably robust, compact and illumination invariant [4, 10].

2.3 Support Vector Machine (SVM)

The basic form of SVM classifier can be expressed as:

(3)

Where input vector **x** e R, w is a normal vector of separating hyper-plane in the feature space produced from mapping of a function (can be linear or non-linear, n can be finite or infinite), and b is a bias. The sign of j(**x**) tells vector **x** belongs to class 1 or class -1.

By solving the QP optimization problem for SVM [8], we have the formal expression of SVM classifier:

$$j(x) = \sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b \qquad (4)$$

Where K is a kernel function: $K(x_i, x) = \phi(x_i) \cdot \phi(x_i)$. By the kernel function, it is not necessary to know the explicit form of $\phi(x)$. Each training sample is associated with a Lagrange coefficient , which is non-zero for Support Vectors (SV).

To enhance the SVM evaluation speed, we should refer back to equation (4), from which we can re-write several possible systematic (formal) ways. One of them is to reduce the number of SVs directly. W is described by a linear combination of SVs and to obtain j(x), x needs to do inner product with all SVs. Thus, reducing the number of SVs can directly reduce the computational cost of SVM in test phase. We refer to works from [11] as an example of successful attempts for reducing SVs, while maintaining acceptable performance of SVM. Another possible way to speed-up SVM test phase is relied on reducing the size of each feature. Reducing the dimension of a feature, while preserving its essential information is one of the usefulness of PCA. By combining these two efforts along SVM, we can get a significant result on reducing the running time of SVM test phase. In this paper, we only adapt the PCA as the speed-up term since RFE considerably reduces the generalization of SVM hyper-plane.

2.4 SVM evaluation with PCA

The motivation to use PCA is to preprocess features by reducing its size while maximizing feature variance before we use SVM. SVM involves inner product spanned by the feature vectors. This is represented by kernel function in eq. 4. The longer the features, more time is needed. Thus PCA is used to speed-up SVM process by reducing features while simultaneously retaining representative parts. This also reduces memory requirements for SVM training. However, projected features by PCA will change the result of SVM since basically SVM just sees another features set. It has been shown in [11] that SVM is invariant under PCA transform thus the properties of SVM are retained. To obtain PCA from our data sets, we use both positives and negatives training images since PCA does not handle labeled data. If we only used positives images, then we will obtain positive data’ principal component only while in testing stage, we do not know whether the new observed data is positive or negative.

3 Experiments

The main tasks of the experiments are to: 1) evaluate the performance of HOG-family on classifying human in datasets. While we evaluate this HOG-family, we enhance the speed of evaluation by PCA. By then we can 2) observe how much speed-up we can achieve without ruining the performance of SVM. The next subsection deals with the chosen performance measure and the datasets we used. The experimental results are presented as Precision-Recall curve to observe the classification quality of SVM trained by HOG-family. An empirical observation also attempted to provide a closer look at the effect of putting PCA on SVM evaluation.

3.1 Performance measure

To quantify the HOG-family performance on binary classifier SVM we use Precision-Recall (PR) curve. We use this measure because we are interested in knowing how many of the objects it detects, and how often the detections it makes are false. Another reason is that most of the majorities of the data are negatives, a small number of true positives will gives small effect to e.g. DET or ROC curve shape, and thus it is difficult to observe the differences in performance between experiment results. PR curve measure the proportion of recall defined as true positive against the proportion of precision that is defined as number of true detected divided by total true hypothesis.

3.2 Dataset

The specific purpose of this evaluation is to choose a suitable implementation of HOG for human detection system in omnidirectional camera mounted on a mobile robot [4]. Thus, we have built a new dataset to represent the application domain of our system. The dataset is taken by our omnidirectional camera system, by then the training images are representative for our purpose. Because the omnidirectional image is converted into panoramic, sampling step degrades the quality of edges and the image contains too much noise and artifacts.

The training data sets contains 249 of 96x160 normalized positives images of person's body as seen in panoramic images in various poses in different illumination condition. We use 98 negatives images which we sampled for training the SVM classifier. Although the dataset has full body size, we only focuses on upper body 48x48 detector since we want the detector to detect human in radius 1-4 m around the robot. If human is too close, then we cannot detect the lower part of the body. The test data sets consist of 100 positives normalized person images and 85 negatives images.

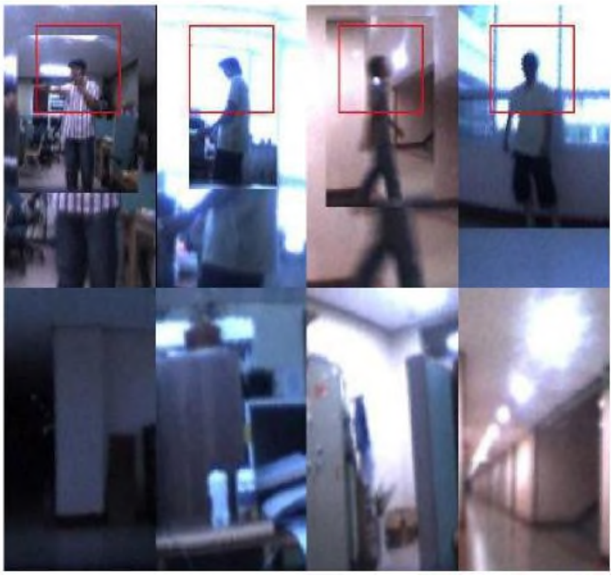


Fig. 2. Examples from our lab person data sets. (Top) positive normalized - the rectangle show 48x48 upper body part - and (Bottom) negative images.

3.3 Experimental results

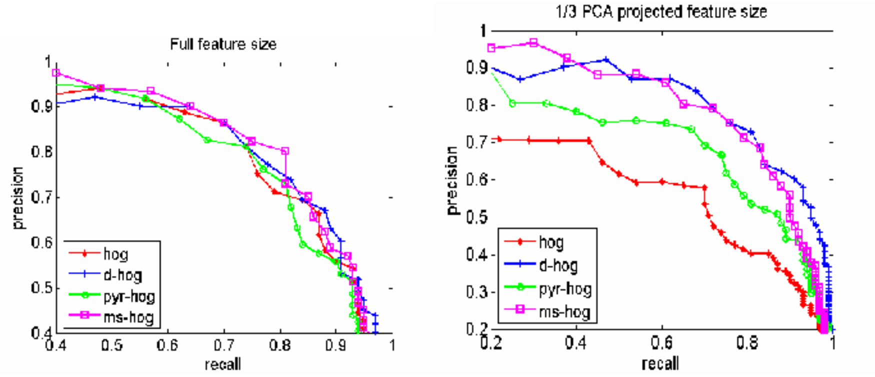
We only use one run phase of training without re-training by augmenting images from false positives obtained with initial training set as used in [8]. The test phase is performed on normalized positives images. Although this might seems restrictive, we may rely on an assumption that given the generalization of the SVM and proper choice of training data sets, if a feature performs well in this data set then it will perform comparably well in the multi-scale detection schema as used in practice [4,8]. The evaluation will be performed under some fixed HOG parameters. For our data set (detection of 48x48 upper body), each cell's size is 12x12 pixels, blocks are defined by 2x2 cells with stride is 12 pixels in both direction (for dense type). We use 249 positives and 953 negatives images for training. The testing data set contain 100 positives and 7332 negatives. To evaluate performances of PCA-SVM we use two projection sizes: one-third and one-sixth of original feature size. This is shown in Table 1. The SVM training and test are performed using svmlight with default parameters and RBF as the kernel.

Table 1. Full and PCA projected size of the feature sets

HOG-type	Full size	1/3	1/6
Non-overlap	144	48	24
Dense	324	108	54
Spatial pyramid	189	63	31
Multi-scale	288	96	48

The performances of HOG-family in PR curves can be seen in Figure 3. For full-size feature, at 0.8 precision, multiscale-HOG which we propose has better recall compared to other HOG variant including dense-HOG from [4,8]. We also can see that PCA reduces performance significantly in 1/6 feature size and for 1/3 feature size, dense-HOG and multi-scale-HOG have comparable performance compared to full feature size. To make this clear, Fig. 4 shows the comparison of full feature size and 1/3 PCA projected feature for dense HOG and multi-scale HOG. To show the advantage of PCA-SVM the processing time to extract feature and the runtime of the test phase for test data set are shown in Table 2. We only consider dense and multiscale since the performance of non-overlap and pyramid on PCA projected feature is much worse than original full-size.

By combining with PCA, at 0.8 precision, recall of multiscale-HOG has dropped nearly 10% while for dense-HOG the recall has dropped approximately 4%. From Table 2, PCA can reduce processing time, however, at the expense of reduces performance as we can see in Figure 3 and Figure 4.



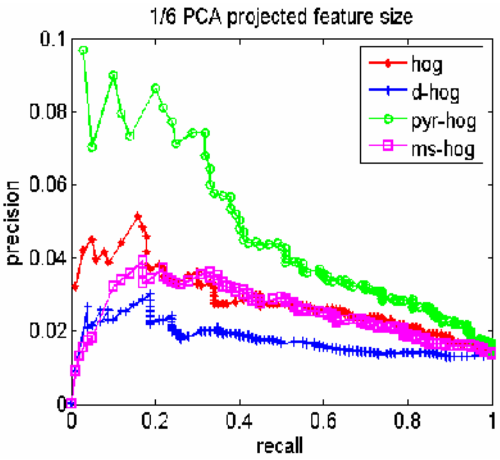


Fig. 3. Performances on Lab data set with RBF kernel. (Top) full feature size (middle) 1/3 PCA projected feature and (bottom) 1/6 PCA projected feature

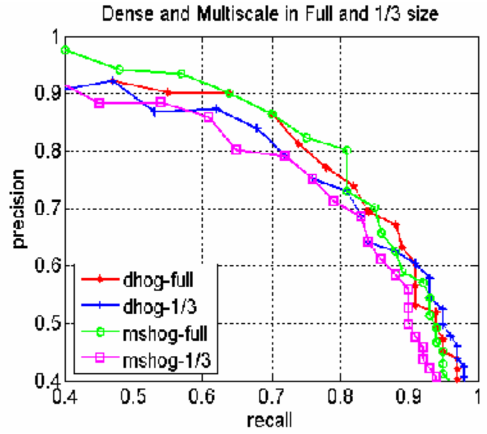


Fig. 4. Performances of full feature size and 1/3 PCA projected feature size for dense and multi-scale HOG.

Table 2. Full and PCA projected size of the feature sets

HOG-type	Feature Size	Feature extraction time(ms)	SVM test(s)
Dense-full	324	23911.69	1.75
Dense-1/3	108	18128.58	0.88
Multiscale-full	288	30830.73	1.68
Uultiscale-1/3	96	26992.65	0.88

3.4 Discussion

Thus, any spatial advantage of feature vectors in unobservable since SVM produce hyper-plane from points in kernel space. Understanding which component of features favorable in classification or retain the spatial configuration can be advantageous in human detection task. We are planning to investigate this in the future. Another important point from this attempt is explaining how the performance of PCA-SVM for human detection is affected by the feature chosen and the data sets. It also depends on size of projection that we choose. Generally, PCA worsen the performance so using PCA might not give any gain even if we can reduce processing time given a challenging data set. If the performance penalty is acceptable then using PCA is preferable. We infer that there is an optimal projection size where PCA can give comparable performance to original size. We also should note that PCA-SVM used in current research is not a principled or integrated approach to reduce feature size because SVM change the feature point in kernel space. That is, no guarantee the new point is the same as original point. This makes the performance penalty unpredictable. We will explore Joint classifier and feature optimization (JCFO which seeks sparsity in its use of both basis functions and features.

4 Conclusion and future direction

This paper has presented an extensive evaluation of HOGfamily features, including the newly proposed: multi-scale HOG, for human detection task. We have assessed the performance of each feature with SVMs classifier and showed the effect of applying PCA for reducing features. Our novel multi-scale image HOG shows best performance compared to well-known dense HOG type in RBF kernel. More integrated combination of reconstructive capability of PCA into SVM machine and multi-scale detection will be carried out in the future.

Acknowledgement. This paper was created with support of Severoceske doly a.s.

SD Pomáháme

References

1. Collins. C, Lipton. A, Kanade. T.: Special Issue on Video Surveillance and Monitoring. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, pp.745-746. 2000.
2. Pentland, A.: Looking at People: Sensing for Ubiquitous and Wearable Computing," IEEE Transaction Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, pp.107-119, Jan, 2000.

3. Gavrila. D. M., Philomin. V.: Real-Time Object Detection for ‘'Smart’' Vehicles. Proceeding IEEE International conference, Computer Vision, pp. 87-93, Sep, 1999.
4. Dalal. N. Triggs, "Histograms of oriented gradients for human detection," Proceeding International Conference on Computer Vision & Pattern Recognition, pp. 886-893, 2005.
5. Tuzel. O, Porikli, F., and Meer. P.: Human detection via classification on Riemannian manifolds. Proc of the 2007, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007,
6. Dalal. N.: Finding people in images and videos, PhD Thesis, Institute National Polytechnique de Grenoble (2006).
7. Lei, H., Govindaraju, V.: Speeding up Multi-class SVM by PCA and Feature Selection". Proc. Feature Selection in Data Mining (FSDM05), The 5th SIAM International Conference on Data Mining Workshop, California, USA, 2005.
8. Freeman, W.T., Roth, M.: Orientation histogram for hand gesture recognition. Proceeding International Workshop on Automatic Face and Gesture Recognition, pp. 296-301, 1995.
9. Lowe. D.G.: Distinctive image features from scale invariant key points," International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.
10. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. Proceeding 6th ACM international conference on Image and video retrieval, pp. 401-408, 2007.
11. Downs, T., Gates, K., Masters, A.: Exact simplification of support vector solutions. Journal of Machine Learning Research, Vol. 2, pp. 293-297, 2001.